Supplementary Tables and Figures

Supplementary Table 1. Description of study design for the 34 studies participating in the Breast Cancer Association Consortium (BCAC)

Study, First author, Year (Reference)	Country	Study design	Definition of case patients	Definition of control subjects	Reported participation rates	No. of Case patients and control subjects	Age at diagnosis, y	Ethnicity
Australian Breast Cancer Family Study (ABCFS), Dite, 2003 [1]	Australia	Population-based	All case patients diagnosed < age 40 plus a random sample of those diagnosed ages 40–59 from cancer registries in Victoria and New South Wales, plus a limited number diagnosed aged 60–69; case patients living in Melbourne recruited from 1992–99 and in Sydney from 1993-98.	Identified from the electoral rolls in Melbourne from 1992–98 and Sydney from 1993–99. Frequency matched to case patients by age in 5 year categories.	75% of case patients and 68% of control subjects completed questionnaires.	1610 1077	23–69	European
Amsterdam Breast Cancer Study (ABCS), Schmidt, 2007 [2]	Netherlands	Mixed	All case patients aged <50 and diagnosed from 1974–1994 in 4 Dutch hospitals.	Random women <50 years of age at baseline from 2 population-based prospective studies run by National Institute for Public Health and the Environment, The Netherlands.	85% of case patients and 50% of control subjects for DNA collection	1481 1140	23–50	European
Bavarian Breast Cancer Case patients and Control subjects (BBCC), Fasching, 2008; [3-4]	Germany	Mixed	Consecutive, unselected case patients with invasive breast cancer recruited at the University Breast Centre, Franconia in Northern Bavaria during 2002–2006.	Healthy women with no diagnosis of cancer aged ≥50 years in Northern Bavaria, who were recruited during 2002–2006	95% of case patients and 99% of control subjects provided questionnaire.	1374 1100	22–96	European
Breast Cancer in Galway Genetic Study (BIGGS), Colleran, 2009 [5-6]	Ireland	Hospital-based	Unselected case patients recruited from West of Ireland since 2001. Case patients were recruited from University College Hospital Galway and surrounding hospitals	Women > 60 years with no personal history of any cancer were identified from retirement groups in the West of Ireland during the period 2001–2008.	Not recorded	975 913	24–90	European

Copenhagen General Population Study (CGPS), Bojesen, 2005 [7-8]	Denmark	Hospital-based	Consecutive, incident case patients from 1 hospital with centralized care for a population of 400,000 women from 2001 to the present.	Community control subjects with no history of breast cancer were identified from the Copenhagen General Population Study recruited 2003–2007.	96% of case patients and 46% of control subjects were interviewed and provided a blood sample.	3306 12534	26–100	European
Spanish National Cancer Centre Breast Cancer Study (CNIO-BCS), Milne, 2006 [9]	Spain	Mixed	Two groups of case patients:1) 574 consecutive breast cancer patients, unselected for family history, from 3 public hospitals, 2 in Madrid and one in Oviedo, from 2000 to 2005. 2) 291 case patients with at least one first degree relative also affected with breast cancer, recruited through the CNIO family cancer clinic in Madrid from 2000 to 2004.	Women attending the Menopause Research Centre between 2000 and 2004 and female members of the College of Lawyers attending a free, targeted medical check-up in 2005, all free of breast cancer and all in Madrid	Not recorded.	1105 1249	23–86	European
Gene Environment Interaction and Breast Cancer in Germany (GENICA), Pesch, 2005 [10-11]	Germany	Population-based	Incident breast cancer case patients enrolled between 2000 and 2004 from the Greater Bonn area (by all of the hospitals within the study region); all enrolled within 6 months of diagnosis	Selected from population registries from 31 communities in the greater Bonn area; matched to case patients in 5-year age classes between 2001 and 2004	Response rate 88% for case patients and 67% for control subjects.	1021 1015	23–80	European
Genetic Epidemiology Study of Breast Cancer by Age 50 (GESBC), Chang-Claude, 2000 [12]	Germany	Population-based	All incident case patients diagnosed <50 years of age in 1992–5 in two regions: Rhein-Neckar-Odenwald and Freiburg, by surveying the 38 clinics serving these regions	Selected from random lists of residents of the study regions supplied by population registries; two control subjects were selected for each case, matched by age and study region. Recruitment was carried out 1992–1998.	70.2% of case patients and 61.2% of control subjects completed the questionnaire.	650 1381	24–50	European
Hannover Breast Cancer Study (HABCS), Dork, 2001 [13]	Germany	Mixed	Case patients who received radiotherapy for breast cancer at Hannover Medical School between 1997–2003, unselected for age or family history	Anonymous female blood bank donors at Hannover Medical School, collected from 8/2005–12/2005, with known age and ethnic background	Approx. 80% of case patients and 70% of control subjects contacted agreed to give a blood sample	1108 1015	25–91	European

Helsinki Breast Cancer Study (HEBCS), Syrjakoski, 2000 [14-15]	Finland	Mixed	(1) Consecutive case patients (883) from the Department of Oncology, Helsinki University Central Hospital 1997–8 and 2000, (2) Consecutive case patients (986) from the Department of Surgery, Helsinki University Central Hospital 2001 – 2004, (3) Familial breast cancer patients (536) from the Helsinki University Central Hospital, Departments of Oncology and Clinical Genetics (1995–)	Healthy females from the same geographical region in Southern Finland in 2003.	(1) 79% of all case patients for the 1. consecutive series, (2) 87% of all case patients for the 2. consecutive series, (3) about 90% of the familial case patients. Control subjects (100%).	2247 1287	22–96	European
Karolinska Breast Cancer Study (KARBAC), Lindblom, 1992 [16-17]	Sweden	Mixed	1. Familial case patients from Department of Clinical Genetics, Karolinska University Hospital , Stockholm. 2. Consecutive case patients from Department of Oncology, Huddinge & Söder Hospital, Stockholm 1998–2000	Blood donors of mixed gender from same geographical region.	1. NA 2. 70% of consecutive case patients provided a blood-sample	832 870	24–88	European
Kuopio Breast Cancer Project (KBCP), Hartikainen, 2005 [18-19]	Finland	Hospital-based	Women seen at Kuopio University Hospital between 1990 and 1995 because of breast lump, mammographic abnormality, or other breast symptom who were found to have breast cancer	Age and long-term area-of-residence matched control subjects selected from the National Population Register and interviewed in parallel with the case patients	Case patients: 98% of those contacted; which is 86% of those potentially eligible. Rate among control subjects was not recorded.	492 532	23–92	European
Kathleen Cuningham Foundation Consortium for research into Familial Breast Cancer/Australian Ovarian Cancer Study (KConFab/AOCS), Beesley, 2007[20-21]	Australia and New Zeland	Mixed	Case patients were from multiple-case breast and breast–ovarian families recruited though family cancer clinics from across Australia and New Zealand from 1998 to the present. Case patients were selected for inclusion in BCAC studies if (i) family was negative for mutations in BRCA1 and BRCA2 (ii) youngest case in the family.	Female control subjects were ascertained by the Australian Ovarian Cancer Study identified from the electoral rolls from all over Australia from 2002–2006.	64% of female family members provided questionnaire data	344 1009	19–78	European

Leuven Multidisciplinary Breast Centre (LMBC), De Maeyer, 2008 [22-23]	Belgium	Mixed	All patients diagnosed with breast cancer and seen in the Multidisciplinary Breast Center in Leuven (Gashuisberg) since June 2007 plus retrospective collection of case patients diagnosed since 2000	Healthy control subjects (blood donors) collected at the Red Cross located in Gasthuisberg hospital (Oct-2007–March 2008)	At least 90% of new patients diagnosed and control subjects agreed to participate in the study.	1206 1142	19–89	European
Mammary Carcinoma Risk Factor Investigation (MARIE), Flesch-Janys, 2008 [24]	Germany	Population-based	Incident and prevalent case patients diagnosed from 2001–2005 in Hamburg in Northern Germany, and from 2002–2005 in Rhein-Neckar-Karlsruhe in Southern Germany.	2 control subjects per case were randomly drawn from population registries and frequency matched by birth year and study region to the case. Control subjects were recruited from 2002 to 2006.	64.1% of case patients & 43.4% of control subjects provided questionnaire data.	3580 7341	50–74	European
Milan Breast Cancer Study Group (MBCSG), Catucci, 2009 [25-26]	Italy	Mixed	Familial and/or early onset breast cancer patients (aged 22–87) negative for mutations in BRCA genes, ascertained in two large cancer centres in Milan from 2000 to date.	Healthy blood donors aged 18–71 years, retruited at two blood centres in Milan from 2004 (centre 1) and 2007 (centre 2) to date	>99%	277 1243	21–80	European
Mayo Clinic Breast Cancer Study (MCBCS), Olson, 2007 [27]	USA	Mixed	Incident case patients residing in 6 states (MN, WI, IA, IL, ND, SD) seen at the Mayo Clinic in Rochester, MN from 2002–5	Women without cancer presenting for general medical examination at the Mayo Clinic. Control subjects were recruited concurrently with case patients and were frequency matched to case patients on age, ethnicity and county/state	68% for case patients, 77% for control subjects were interviewed and provided a blood sample	1202 1574	22–89	European
Melbourne Collaborative Cohort Study (MCCS), Giles, 2002 [28]	Australia	Prospective cohort	Incident case patients diagnosed within the Melbourne Collaborative Cohort Study during the follow-up from baseline (1990–1994) to 2004 of the 24469 participating women	Random sample of the initial cohort	All incident case patients and all the control subjects in the random sample.	1234 778	30–82	European
Multiethnic Cohort (MEC), Kolonel, 2000 [29]	USA	Prospective cohort	Incident case patients identified from SEER cancer registries in Los Angeles County & State registries in California & Hawaii, USA from 1993–2002. Grouped by self-reported ethnicity.	Women without cancer from the same States, recruited concurrently with case patients & frequency matched to case patients by age at blood-draw & self-reported ethnicity.	>60% for both case patients & control subjects	873 829	46–82	European (52%) Asian (48%)

Northern California Breast Cancer Family Registry (NC-BCFR), John, 2004 [30]	USA	Population-based	Incident case patients aged <65 years diagnosed between 1995 and 2003 were identified through the SEER cancer registry of the Greater Bay Area Cancer Registry. Enrolled all case patients meeting NC-BCFR criteria (dx at age <35 yrs, personal history of ovarian or childhood cancer, bilateral breast cancer with 1st dx at age <50, family history of breast or ovarian cancer in first-degree relatives) and a random sample of case patients not meeting the NC-BCFR criteria.	Control subjects were identified through random digit dialing conducted from 1999–2000 in the same geographic region. Control subjects were frequency matched to case patients on 5-year age group and race/ethnicity, at a ratio of 1 control per 2 case patients diagnosed from 1995–1998.	Case patients: Response to telephone screening 86%; to in-person interview and blood: 60%. Control subjects: response to in-person interview and blood 50%.	1399 337	51–64	all
Nurses Health Study (NHS), Hankinson, 1998 [31-32]	USA	Prospective cohort	Incident case patients arising in the sub-cohort of 32,826 cohort members who gave a blood specimen in 1989–1990 are included if they were diagnosed with breast cancer prior to July 1, 2000.	Control subjects were women in this sub-cohort who were not diagnosed with breast cancer.	All incident case patients and selected control subjects are included.	1029 1761	44–79	European
Oulu Breast Cancer Study (OBCS), Erkko, 2007[33]	Finland	Mixed	Consecutive incident case patients diagnosed at the Oulu University Hospital between 2000 and 2004.	Healthy, consecutive, anonymous, female Finnish Red–Cross blood donors recruited in 2002 from the same geographical region in Northern Finland.	All of the asked control subjects, and 71% of all case patients.	537 511	28–92	European

Ontario Familial Breast Cancer Registry (OFBCR), John, 2004 [30]	Canada	Population-based	Case patients diagnosed between 1 Jan 1996–31 Dec 1998 were identified from the Ontario Cancer Registry. All women with invasive breast cancer aged 20–54 years who met the OFBCR definition for high genetic risk (family history of specific cancers particularly breast and ovarian, early onset disease, Ashkenazi ethnicity or a diagnosis of multiple breast cancer), a 25% random sample of individuals in this age category who did not meet the OFBCR definition, 35% of those aged 55–69 at high risk and 8.75% aged 55–69 at low risk were asked to participate.	Unrelated, unaffected population control subjects were recruited between 2003–2005 by calling randomly selected residential telephone numbers throughout the same geographical region. Eligible control subjects were women with no history of breast cancer and were frequency-matched by 5-year age group to the expected age distribution of case patients.	Case patients: consent to contact patients was 92%, response to initial family history questionnaire was 65%, response to risk factor questionnaires was 73% of all eligible. Control subjects: approximately, 65% of identified eligible women returned questionnaires.	1407 367	22–81	
Leiden University Medical Centre Breast Cancer Study (ORIGO), de Bock, 2004 [34-35]	Netherlands	Mixed	Consecutive case patients diagnosed 1996–2006 in 2 hospitals of South–West Netherlands (Leiden & Rotterdam). No selection for family history; Rotterdam case patients selected for diagnosis aged <70. Case patients with in situ carcinomas eligible.	Three groups of control subjects: (1) Blood bank healthy donors from Southwest Netherlands recruited in 1996, 2000 or 2007; (2) People who married a person who was part of a family with high breast cancer risk (BRCA1/2/x). From the Southwest of the Netherlands, recruited 1990–1996; (3) Females tested at the local clinical genetics department for familial diseases, excluding familial cancer syndromes (no mutation found in gene(s) related to the disease being tested), recruited 1995–2007.	80–90%	1326 1663	22–88	European

NCI Polish Breast Cancer Study (PBCS), Garcia-Closas, 2006 [36]	Poland	Population-based	Incident case patients from 2000–2003 identified through a rapid identification system in participating hospitals covering ~ 90% of all eligible case patients; periodic check against the cancer registries in Warsaw and Łódź to assure complete identification of case patients	Randomly selected from population lists of all residents of Poland, stratified and frequency matched to case patients by case city and age in 5 year categories. Recruited 2000–2003.	79% of eligible case patients and 69% of eligible control subjects agreed to personal interview.	2000 2378	27–75	European
Prospective Study of Outcomes in Sporadic Versus Hereditary Breast Cancer (POSH), Eccles, 2007 [37-38]	UK	Population-based	Case patients aged 40 or younger at breast cancer diagnosis. Recruited across UK and diagnosed between January 2000 to December 2007	No in-house control subjects	DNA available for 95% of participants	1001 0	18–40	3% ethnic minority groups eg Black, Asian
Rotterdam Breast Cancer Study (RBCS), Easton, 2007 [39]	Netherlands	Hospital-based	Familial breast cancer patients selected from the clinical genetics center at Erasmus Medical Center; recruited 1994 – 2005	Spouses or mutation–negative siblings of heterozygous Cystic Fibrosis mutation carriers selected from the clinical genetics center at Erasmus Medical Center; recruited 1996 – 2006	100% of case patients and control subjects provided a blood sample.	747 801	18–84	European
Singapore and Sweden Breast Cancer Study (SASBAC), Wedren, 2004 [40]	Sweden	Population-based	Incident case patients from October 1993 to March 1995 identified via the 6 regional cancer registries in Sweden, to which reporting is mandatory.	Control subjects were randomly selected from the total population registry in 5-year age groups to match the expected age-frequency distribution among case patients. Patients and control subjects were recruited from Oct 1993 through April 1995.	84% of case patients & 82% of control subjects completed questionnaire.	1701 1524	50–75	European
Sheffield Breast Cancer Study (SBCS), MacPherson, 2004 [41-42]	UK	Mixed	Women with pathologically confirmed breast cancer recruited from surgical outpatient clinics at the Royal Hallamshire Hospital, Sheffield, 1998 – 2002; case patients are a mixture of prevalent and incident disease	Unselected women attending the Sheffield Mammography Screening Service between Sep 2000 – Aug 2002, if their mammograms showed no evidence of a breast lesion	Not recorded	1115 1271	29–93	European

Study of Epidemiology and Risk factors in Cancer Heredity (SEARCH), Lesueru, 2005 [43]	UK	Mixed	2 groups of case patients identified through East Anglian Cancer Registry; 1) prevalent case patients diagnosed age <55 from 1991–6 and alive when study started in 1996; 2) incident case patients diagnosed age < 70 diagnosed after 1996	Two groups of control subjects: (1) selected from the EPIC–Norfolk cohort study of 25,000 individuals age 45–74 recruited between 1992 and1994, based in the same geographic region as case patients; (2) selected from GP practices from March 2003 to present, frequency matched to case patients by age and geographic region	64% of eligible case patients and 41% of invited control subjects provided a blood sample	6882 8096	23–69	European
IHCC–Szczecin Breast Cancer Study (SZBCS), Jakubowska, 2009 [44-45]	Poland	Mixed	Prospectively ascertained case patients of invasive breast cancer patients diagnosed at the Regional Oncology Hospital (Szczecin) in the years 2002, 2003, 2006 and 2007 or the University Hospital from 2002 to 2007 in Szczecin, West–Pomerania, Poland.	Unaffected, matched to case patients for year of birth, sex and region; from families with negative cancer family history; control subjects were part of a population–based study of the 1.3 million inhabitants of West Pomerania performed in 2003 and 2004 designed to identify familial aggregations of cancer by our centre	>95% case patients and 55% control subjects	807 1032	26–88	European
IARC–Thai Breast Cancer Study (TBCS), Sangrajrang, 2008 [46]	Thailand	Hospital–based	Incident case patients diagnosed at the National Cancer Institute (NCI) in Bangkok and Khon Kaen Hospital during the period May 2002–March 2004.	Control subjects were randomly selected healthy females visiting hospital patients with diseases other than breast or ovarian cancer at NCI Bangkok and Khon Kaen Hospital during the period May 2002–March 2004.	94% of case patients and 73% of control subjects completed a questionnaire.	474 390	17–81	Asian
Taiwanese Breast Cancer Study (TWBCS), Ding, 2009 [47-48]	Taiwan	Hospital–based	Incident case patients diagnosed & treated at 2 major teaching hospitals in Taiwan between March 2002 and August 2005.	Control subjects were cancer–free individuals, randomly selected from women attending health exam at same hospital during study period. Underwent 1–day health examination – any showing evidence cancer excluded.	>90% case patients & ~ 40% of control subjects	909 1410	18–87	Asian

UCI Breast Cancer Study (UCIBCS), Anton–Culver, 2000 [49-50]	USA	Population–based	All case patients diagnosed in Orange County, California, during one–year period beginning March 1, 1994. Ascertained through the population–based Cancer Surveillance Program of Orange County California (CSPOC).	Female control subjects under age 75 years without history of cancer recruited using random digit dialing among Orange County residents & frequency matched to case patients by age & race/ethnicity. Recruited from 1998–2003	Case patients 76% and Control subjects 80%	933 633	24–90	European Asian Hispanic

Supplementary Table 2. Number of breast cancer case patients with risk factor data in 34 BCAC studies

Study	Age at Menarche	Parity	Age at First birth	Family History	Body mass index, kg/m^2
ABCFS	1354	1133	1057	1360	1352
ABCS	0	0	0	493	0
BBCC	933	1024	819	1023	1014
BIGGS	0	0	0	259	0
CGPS	1106	1369	289	1870	331
CNIO–BCS	216	134	0	0	0
GENICA	971	972	798	972	972
GESBC	525	527	408	527	527
HABCS	416	650	0	766	0
HEBCS	1575	1648	1183	2147	1516
KARBAC	379	451	339	450	0
KBCP	422	437	349	437	426
KConFab/AOCS	191	191	173	170	189
LMBC	551	680	191	694	492
MARIE	2274	2553	2123	2492	2550
MBCSG	34	34	22	34	34
MCBCS	932	1042	895	977	999
MCCS	877	879	714	879	879
MEC	0	781	0	790	781
NC–BCFR	1195	1226	960	1226	1224
NHS	0	903	0	914	0
OBCS	435	455	0	310	0
OFBCR	859	746	714	1000	838
ORIGO	720	750	731	977	938
PBCS	1795	1808	1542	1808	1808
POSH	0	0	0	988	959
RBCS	148	546	411	548	33
SASBAC	956	1058	907	1034	1053
SBCS	715	715	591	723	691

SEARCH	4244	4428	3697	4494	4406
SZBCS	0	0	0	754	0
TBCS	0	243	170	243	242
TWBCS	729	733	661	755	736
UCIBCS	754	753	629	754	689
Total	**25306**	**28869**	**20373**	**32868**	**25679**

Supplementary Table 3. Marker assessment methods and definitions of staining positivity for studies with data available

Study*	Antibody	Vendor, location	Clone	Source†	Definition of positive stain
ABCS	Monoclonal, Mouse anti-human ER	Neomarkers, Labvision, Fremont, CA	1D5 and 6F11	T	>10% cells stained
GENICA	Monoclonal, Mouse anti-human ER	DAKO, Carpinteria, CA	1D5	S	Number of cells x intensity (german immuno reactive score) 3–12 = positive
HEBCS	Monoclonal, Mouse anti-human ER	Novocastra, Newcastle upon Tyne, UK			>10% cells stained
KBCP	Monoclonal, Mouse anti-human ER	Abbot Laboratories, Abbot Park, IL	ER_ICA kit	S and R	Intensity score (0.1,2,3)*percentage score (0,1,2,3)= 3–6 (score range 0–6)
MARIE	Monoclonal, Mouse anti-human ER	DAKO, Carpinteria, CA	ID5	R	≥10% tumour nuclei stained with intensity score (0,1,2,3) > 1
MCBCS	Monoclonal, Mouse anti-human ER	Novocastra, Newcastle upon Tyne, UK	6F11/2	S	Any nuclear staining
MCCS	Monoclonal, Rabbit anti-human ER	NeoMarkers, Fremont, CA	SP1 clone	S and R	Nuclei positive with intensity score (0,1,2,3) >=1
PBCS	Monoclonal, Mouse anti-human ER	Novocastra, Newcastle upon Tyne, UK	6F11/2 (1D5 for AQUA)	T and R	Intensity score (0,1,2,3) * percentage of cells stained (0–100%) >=10 (total score range 0–300)
RBCS	Monoclonal, Rabbit anti-human ER	Thermo Scientific, Fremont, CA	SP1	R	>10% cells stained
SBCS	Monoclonal, Mouse anti-	Vector laboratories,	6F11/2	T and R	Intensity score (0,1,2,3) * percentage of cells stained (0–100%) >=50 (total score range 0–300)

	human ER	Burlingame, CA			
SEARCH	Monoclonal, Mouse anti-human ER	Novocastra, Newcastle upon Tyne, UK	6F11/2	T and R	Allred score (intensity*percentage)= 3–8 (score range 0–8)
ABCS	Monoclonal, Mouse anti-human PR	ImmunoLogic, Duiven, The Netherland	PR-1	T	>10% cells stained
GENICA	Monoclonal, Mouse anti-human PR	Dako DAKO, Carpinteria, CA	PgR 636	S	Number of cells x intensity (german immuno reactive score) 3–12 = positive
HEBCS	Monoclonal, Mouse anti-human PR	Dako DAKO, Carpinteria, CA			>10% cells stained
KBCP	Monoclonal, Rat anti-human ER	Abbot Laboratories, Abbot Park, IL	PR–ICA kit	S and R	Intensity score (0.1,2,3)*percentage score (0,1,2,3)= 3–6 (score range 0–6)
MARIE	Monoclonal, Mouse anti-human PR	Dako DAKO, Carpinteria, CA	PgR 636	R	≥10% tumour nuclei stained with intensity score (0,1,2,3) > 1
MCBCS	Monoclonal, Mouse anti-human PR	Dako DAKO, Carpinteria, CA	PgR 636	S	Any nuclear staining
MCCS	Monoclonal, Mouse anti-human PR	Dako DAKO, Carpinteria, CA	PgR 636	S and R	Nuclei positive with intensity score (0,1,2,3) >=1
PBCS	Monoclonal, Mouse anti-human PR	DakoCytomation, Glostrup, Denmark	PgR 636	T and R	Intensity score (0,1,2,3) * percentage of cells stained (0–100%) >=10 (total score range 0–300)
RBCS	Monoclonal, Mouse anti-human PR	DAKO, Carpinteria, CA	PgR 636	R	>10% cells stained
SBCS	Monoclonal, Mouse anti-human PR	Vector laboratories, Burlingame, CA	1A6	T	Allred score (intensity*percentage)= 3–8 (score range 0–8)
SEARCH	Monoclonal, Mouse anti-human PR	DAKO, Carpinteria, CA	PgR 636	T and R	Allred score (intensity*percentage)= 3–8 (score range 0–8)

ABCS	Monoclonal, Mouse anti-human HER2	NeoMarkers, Fremont, CA	3B5 and 23	T	Score 3+
GENICA	Polyclonal, Rabbit anti-human HER2	DAKO, Carpinteria, CA	HercepTest™	S	Score 2+
HEBCS	Monoclonal, Mouse anti-human HER2/ digoxigenin-labeled HER-2 probe	Novocastra, Newcastle upon Tyne, UK; Zymed Laboratories, South San Francisco, CA	NCL–CB11/HER2 CISH probe	T	CISH result (0–1=neg, 2–3=pos; if no CISH result: IHC 0–1=neg, 3=pos
MARIE	Polyclonal, Rabbit anti-human HER2	DAKO, Carpinteria, CA	A 4085	R	Score 3+ in ≥ 30% stained tumor cells or FISH amplified
MCBCS	Polyclonal, Rabbit anti-human HER2	DAKO, Carpinteria, CA	HercepTest™	S	Complete strong cytoplasmic staining in >30% tumor cells
MCCS	Polyclonal, Rabbit anti-human HER2	DAKO, Carpinteria, CA	A 4085	S	Score 2+
PBCS	Polyclonal, Rabbit anti-human HER2	DAKO, Carpinteria, CA	HercepTest™	T	Score 3+ in ≥20% stained tumor cells
SBCS	Polyclonal, Rabbit anti-human HER2	DAKO, Carpinteria, CA	HercepTest™	T	Score 2+
SEARCH	Polyclonal, Rabbit anti-human HER2	DAKO, Carpinteria, CA	HercepTest™	T and R	Score 2+
HEBCS	Monoclonal, Mouse anti-	Zymed Laboratories,	31G7	T	Intensity score (0,1,2,3)* percentage of cells stained (0–100%) >10

	human EGFR	South San Francisco, CA			
KBCP	Monoclonal, Mouse anti-human EGFR	NeoMarkers, Fremont, CA	MS-1868-S1	T	Intensity score (0.1,2,3)*percentage score (0,1,2,3)= 3–6 (score range 0–6)
MCBCS	Monoclonal, Mouse anti-human EGFR	DAKO, Carpinteria, CA	2-18C9	S	Any cytomplasmic membrane staining
MCCS	Monoclonal, Mouse anti-human EGFR	Zymed Laboratories, South San Francisco, CA	31G7	S	Intensity score (0,1,2,3)>=1
PBCS	Monoclonal, Mouse anti-human EGFR	Zymed Laboratories, South San Francisco, CA	31G7	T	Intensity score (0,1,2,3)* percentage of cells stained (0–100%) >10
SEARCH	Monoclonal, Mouse anti-human EGFR	Zymed Laboratories, South San Francisco, CA	31G7	T	Allred score (intensity*percentage)= 3–8 (score range 0–8)
ABCS	Monoclonal, Mouse anti-human CK5/6	DAKO, Carpinteria, CA Zymed Laboratories, South San Francisco, CA	D5/16	T	>1% cells stained
HEBCS	Monoclonal, Mouse anti-human CK5/6	DAKO, Carpinteria, CA	M7237	T	>=10% positive
KBCP	Monoclonal, Mouse anti-human CK5/6	DAKO, Carpinteria, CA	M7237	T	Intensity score (0.1,2,3)*percentage score (0,1,2,3)= 3–6 (score range 0–6)

MCBCS	Monoclonal, Mouse anti-human CK5/6	Zymed Laboratories, South San Francisco, CA	D5/16 B4	S	>10% cells stained
MCCS	Monoclonal, Mouse anti-human CK5/6	Zymed Laboratories, South San Francisco, CA	Monoclonal, Mouse anti-human CK5/6	S	Intensity score (0,1,2,3)>=1
PBCS	Monoclonal, Mouse anti-human CK5	Novocastra, Newcastle upon Tyne, UK;	D5/16 B4	T	Intensity score (0,1,2,3)* percentage of cells stained (0–100%) >10
SBCS	Monoclonal, Mouse anti-human CK5/6	Vector laboratories, Burlingame, CA	XM26	T	>10% cells stained
SEARCH	Monoclonal, Mouse anti-human CK5/6	DAKO, Carpinteria, CA	D5/16 B4	T	>10% cells stained

*The following studies obtained ER/PR status from medical records: ABCFS, BBCC, BIGGS, CGPS, CNIO–BCS, GENICA, GESBC, HABCS, HEBCS, KARBAC, kConFab/AOCS, LMBC, MARIE, MBCSG, MEC, NC–BCFR, NHS, ORIGO, POSH, SASBAC, SZBCS, UCIBCS. The following studies obtained HER2 status from medical records: BBCC, CNIO–BCS, GENICA, KBCP, LMBC, MARIE, POSH.

†Source: T= Tissue microarray. S = whole tumor sections. R = Hospital/Pathology/Cancer Registry record.

Supplementary Table 4. Associations between tumor characteristics and breast cancer subtypes defined by ER, PR, and HER2*

Age and tumor characteristics	Tumor subtypes†										
	$ER^+/HER2^-$ or $PR^+/HER2^-$ (N=9,423)		$ER^+/HER2^-$ or $PR^+/HER2^-$ (N=1,621)			$ER^-/PR^-/HER2^+$ (N=937)			$ER^-/PR^-/HER2^-$ (N=1,963)		
	No.	%	No.	%	*P*‡	No.	%	*P*‡	No.	%	*P*‡
Age, y											
<40	611	(6)	314	(19)	referent	149	(16)	referent	314	(16)	referent
40-49	2,174	(23)	391	(24)	.009	222	(23)	.01	494	(25)	6×10^{-8}
50-59	2,822	(30)	412	(25)	.0003	285	(30)	.39	619	(31)	2×10^{-4}
60-69	2,709	(28)	402	(24)	2×10^{-5}	219	(23)	.008	380	(19)	1×10^{-14}
≥70	1,218	(13)	138	(8)	2×10^{-7}	78	(8)	2×10^{-4}	190	(9)	2×10^{-10}
Tumor Grade											
Well differentiated	2,338	(25)	119	(8)	referent	15	(2)	referent	77	(4)	referent
Moderately differentiated	5,347	(56)	791	(50)	4×10^{-21}	242	(26)	1×10^{-11}	465	(23)	4×10^{-14}
Poorly differentiated	1,850	(19)	673	(42)	1×10^{-54}	678	(72)	5×10^{-45}	1,436	(73)	2×10^{-130}
Tumor Histology											
Ductal	5,737	(69)	957	(81)	referent	742	(93)	referent	1,454	(83)	referent
Lobular	1,704	(20)	137	(12)	6×10^{-9}	14	(2)	2×10^{-17}	115	(7)	1×10^{-12}
Medullary	61	(1)	13	(1)	.42	12	(2)	.98	68	(4)	7×10^{-8}
Other	872	(10)	74	(6)	.23	26	(3)	3×10^{-5}	106	(6)	.06
Tumor Size											
0.1 - 1.0	1,170	(18)	139	(14)	referent	73	(12)	referent	168	(12)	referent
1.1 - 2.0	2,959	(45)	376	(38)	0.2	230	(37)	.36	530	(36)	.12
> 2.0	2,393	(37)	471	(48)	0.58	318	(51)	.49	760	(52)	.04
Axillary node status											
Negative	5,494	(63)	805	(55)	referent	422	(48)	referent	1,090	(59)	referent
Positive	3,271	(37)	655	(45)	.25	456	(52)	.002	748	(41)	6×10^{-6}

*Unconditional logistic regression models were used to estimate associations between tumor subtypes and age and tumor characteristics, where tumor subtypes were the outcome variable and tumor characteristics, age at diagnosis, and study were independent variables.

†Defined by expression levels of ER, PR, and HER2 in tumors. Expression data were based on immunohistochemical staining and pathologist readings and/or imaging analysis.

‡*P* values were calculated using two-sided Wald test.

Supplementary Table 5. Associations between number of pregnancies and tumor subtypes in case–case analyses*

Tumor subtypes†	No. of studies	No. of pregnancies							
		1		2		≥3		≥3 vs 1	
		No.	(%)	No.	(%)	No.	(%)	OR (95% CI)	*P*‡
ER^+	30	4,315	(23)	8,160	(44)	6,165	(33)	1.00 (referent)	
ER^-	30	1,338	(23)	2,544	(44)	1,944	33)	1.08 (0.99 to 1.18)	.09
PR^+	30	3,334	(24)	6,102	(43)	4,766	(34)	1.0 (referent)	
PR^-	30	1,843	(24)	3,201	(42)	2,527	(33)	1.01 (0.93 to 1.10)	.81
ER^+/PR^+	30	3,102	(23)	5,701	(43)	4,445	(34)	1.00 (referent)	
ER^+/PR^-	30	783	(26)	1,223	(41)	982	(33)	0.92 (0.82 to 1.04)	.19
ER^-/PR^+	30	216	(24)	385	(43)	303	(34)	1.06 (0.87 to 1.30)	.57
ER^-/PR^-	30	1,053	(23)	1,961	(43)	1,527	(34)	1.08 (0.97 to 1.19)	.16
$ER^+/HER2^-$ or $PR^+/HER2^-$	15	1,858	(28)	2,857	(43)	1,946	(29)	1.00 (referent)	
$ER^+/HER2^+$or $PR^+/HER2^+$	15	282	(28)	436	(44)	283	(28)	0.88 (0.73 to 1.08)	.22
$ER^-/PR^-/HER2^+$	15	178	(29)	256	(41)	188	(30)	1.08 (0.85 to 1.37)	.54
$ER^-/PR^-/HER2^-$	15	356	(26)	586	(43)	432	(31)	1.24 (1.04 to 1.48)	.01

*Unconditional logistic regression models were used to estimate associations between tumor subtypes and age and tumor characteristics, where tumor subtypes were the outcome variable and number of pregnancies, age at diagnosis, and study were independent variables.

†Defined by expression levels of ER, PR, and HER2 in tumors. Expression data were based on immunohistochemical staining and pathologist readings and/or imaging analysis.

‡*P* values were calculated using two-sided Wald test.

Supplementary Table 6. Associations between BMI among women >50 years old with by tumor size and tumor subtypes in case–case analyses*

Tumor subtypes†	No. of studies	BMI among women > 50 years old, kg/m^2							
		<25		25–30		≥30		≥30 vs <25	
		No.	(%)	No.	(%)	No.	(%)	OR (95% CI)	*P*‡
Large tumors (>2 cm)									
ER$^+$	17	1,017	(37)	1,040	(38)	678	(25	1.00 referent)	
ER$^-$	17	369	(34)	408	(38)	305	(28)	1.16 (0.97 to 1.39)	.10
	17								
PR$^+$	17	66	(35)	706	(37)	502	(27)	1.00 (referent)	
PR$^-$	17	550	(38)	550	(38)	368	(25)	0.85 (0.71 to 1.01)	.06
	17								
ER$^+$/PR$^+$	17	623	(36)	663	(38)	462	(26)	1.00 (referent)	
ER$^+$/PR$^-$	17	252	(42)	217	(37)	125	(21)	0.65 (0.51 to 0.84)	.0008
ER$^-$/PR$^+$	17	37	(31)	42	(35)	40	(34)	1.38 (0.88 to 2.16)	.16
ER$^-$/PR$^-$	17	294	(34)	332	(38)	241	(28)	1.03 (0.83 to 1.26)	.80
ER$^+$/HER2$^-$ or PR$^+$/HER2$^-$	10	364	(34)	410	(38)	312	(29)	1.00 (referent)	
ER$^+$/HER2$^+$or PR$^+$/HER2$^+$	10	50	(38)	51	(38)	32	(24)	0.98 (0.61 to 1.56)	.93
ER$^-$/PR$^-$/HER2$^+$	10	50	(41)	49	(40)	23	(19)	0.58 (0.35 to 0.97)	.04
ER$^-$/PR$^-$/HER2$^-$	10	89	(29)	118	(39)	99	(32)	1.20 (0.87 to 1.65)	.26
Small tumors (≤2 cm)									
ER$^+$	17	2,655	(44)	2,197	(36	1,188	(20)	1.00 (referent)	
ER$^-$	17	584	(45)	474	(36)	244	(19)	0.94 (0.80 to 1.11)	.47
	17								
PR$^+$	17	1,793	(43)	1,513	(37)	844	(20)	1.00 (referent)	
PR$^-$	17	1,002	(48)	728	(35)	3 4	(17)	0.81 (0.70 to 0.94)	.006

	17								
ER^+/PR^+	17	1,701	(43)	1,430	(36)	804	(20)	1.00 (referent)	
ER^+/PR^-	17	538	(49)	379	(35)	179	(16)	0.74 (0.61 to 0.89)	.002
ER^-/PR^+	17	86	(43)	78	(39)	36	(18)	0.81 (0.55 to 1.19)	.28
ER^-/PR^-	17	458	(46)	347	(35)	185	(19)	0.90 (0.78 to 1.05)	.26
$ER^+/HER2^-$ or $PR^+/HER2^-$	10	808	(40)	781	(38)	444	(22)	1.00 (referent)	
$ER^+/HER2^+$or $PR^+/HER2^+$	10	81	(46)	66	(37)	31	(17)	0.87 (0.57 to 1.34)	.53
$ER^-/PR^-/HER2^+$	10	56	(45)	46	(37)	22	(18)	0.89 (0.53 to 1.49)	.65
$ER^-/PR^-/HER2^-$	10	126	(43)	98	(34)	66	(23)	0.98 (0.71 to 1.35)	.90

*Unconditional logistic regression models were used to estimate associations between tumor subtypes and age and tumor characteristics, where tumor subtypes were the outcome variable and BMI among women >50 years old, age at diagnosis, and study were independent variables.

†Defined by expression levels of ER, PR, and HER2 in tumors. Expression data were based on immunohistochemical staining and pathologist readings and/or imaging analysis.

‡*P* values were calculated using two-sided Wald test.

Supplementary Table 7: Associations between family history of breast cancer and tumor subtypes in case–case analyses*

Tumor subtypes†	No. of studies	Family history of breast cancer					
		Negative		Positive		Positive vs negative	
		No.	(%)	No.	(%)	OR (95% CI)	*P*‡
ER^+	33	19,711	(79)	5,202	(21)	1.00 (referent)	
ER^-	33	6,354	(81)	1,476	(19)	0.95 (0.89 to 1.02)	.12
PR^+	33	14,609	(79)	3,972	(21)	1.00 (referent)	
PR^-	33	8,038	(80)	1,989	(20)	0.95 (0.89 to 1.01)	.10
ER^+/PR^+	33	13,547	(78)	3,736	(22)	1.00 (referent)	
ER^+/PR^-	33	3,123	(79)	821	(21)	0.95 (0.87 to 1.04)	.28
ER^-/PR^+	33	1,004	(82)	226	(18)	0.96 (0.82 to 1.12)	.57
ER^-/PR^-	33	4,863	(81)	1,163	(19)	0.95 (0.87 to 1.02)	.15
$ER^+/HER2^-$ or $PR^+/HER2^-$	17	7,025	(81)	1,693	(19)	1.00 (referent)	
$ER^+/HER2^+$or $PR^+/HER2^+$	17	1,283	(84)	252	(16)	0.96 (0.82 to 1.12)	.63
$ER^-/PR^-/HER2^+$	17	730	(84)	135	(16)	0.91 (0.74 to 1.11)	.34
$ER^-/PR^-/HER2^-$	17	1,498	(83)	318	(18)	0.99 (0.86 to 1.14)	.89
Analyses restricted to case patients with CBP marker (CK5/6 or CK5 or EGFR) data							
ER^+ or $PR^+/HER2^-$	7	3,422	(79)	938	(22)	1.00 (referent)	
CBP	7	368	(75)	120	(25)	1.38 (1.08 to 1.75)	.01

*Unconditional logistic regression models were used to estimate associations between tumor subtypes and age and tumor characteristics, where tumor subtypes were the outcome variable and family history, age at diagnosis, and study were independent variables.

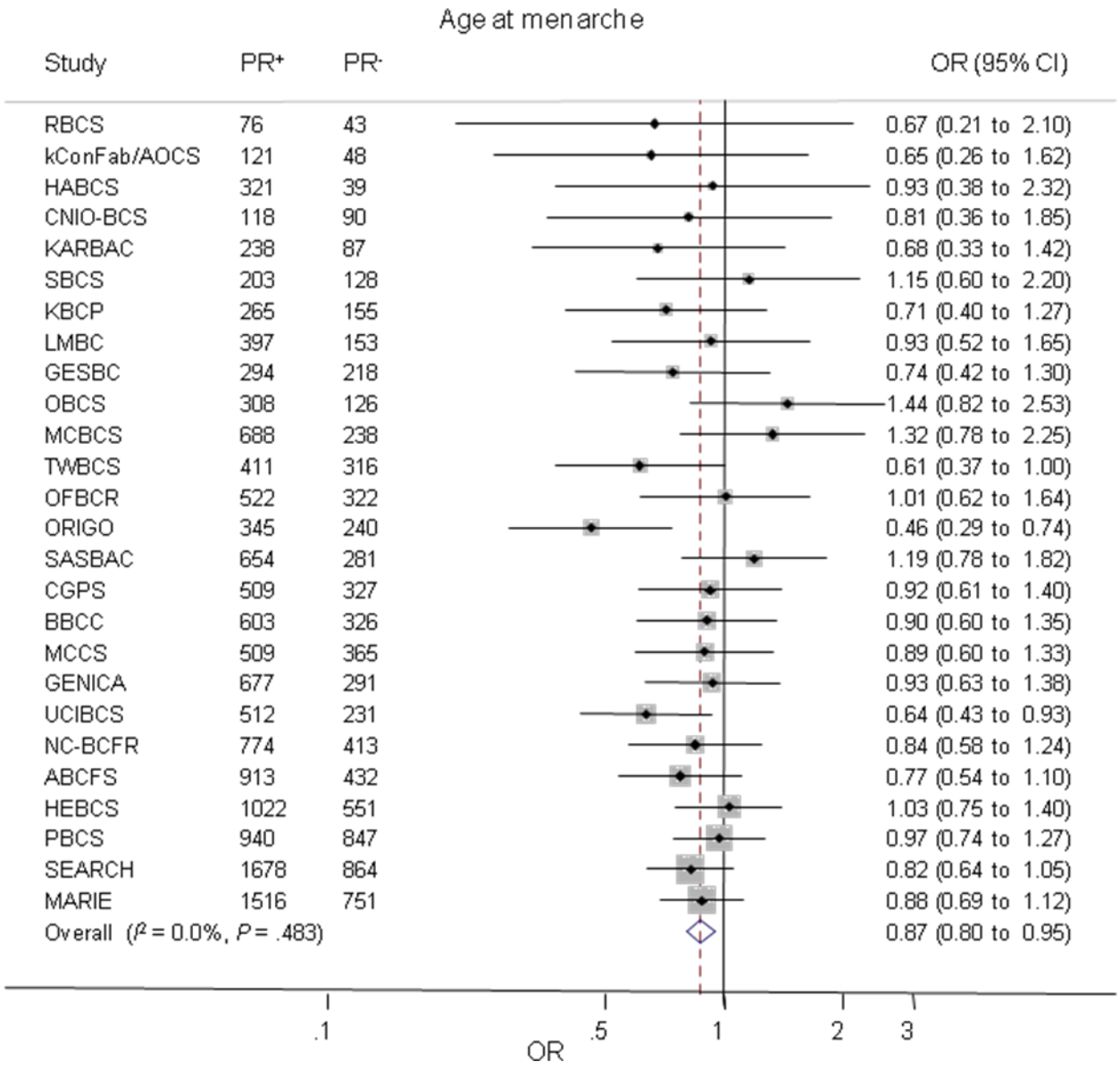
†Defined by expression levels of ER, PR, and HER2 in tumors. Expression data were based on immunohistochemical staining and pathologist readings and/or imaging analysis.

‡*P* values were calculated using two-sided Wald test.

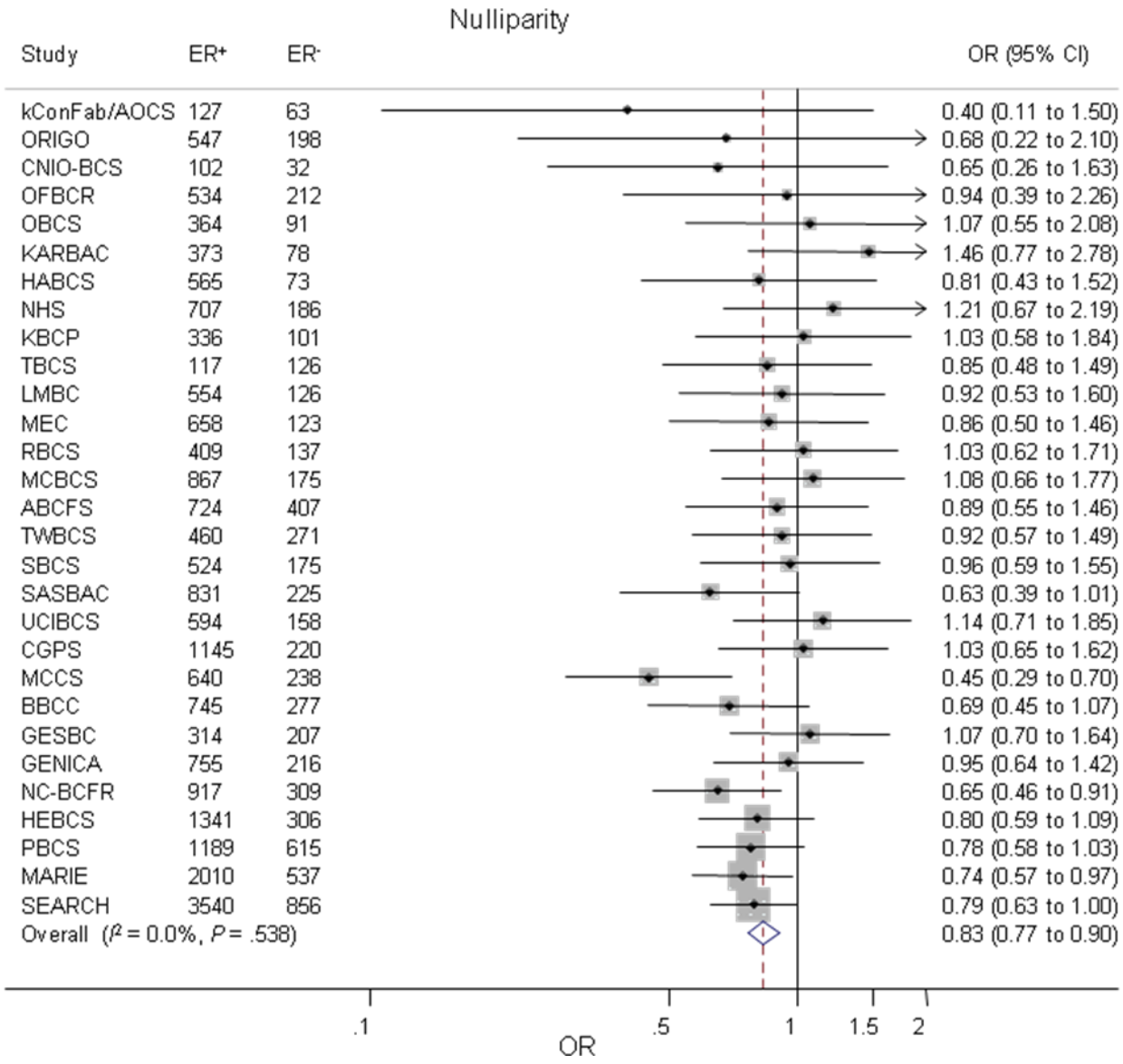
Supplementary Figures

Supplementary Figure 1

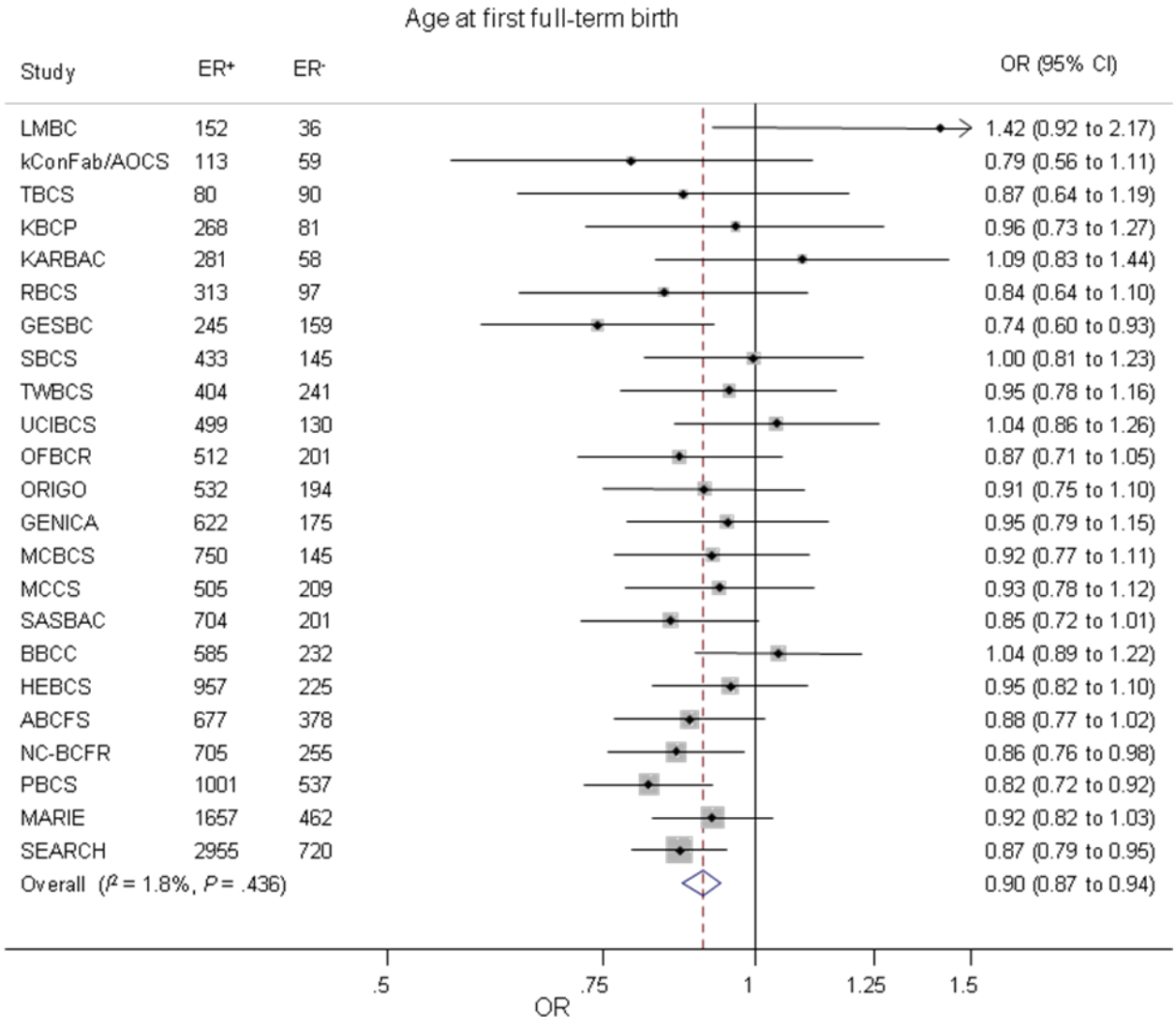
A.



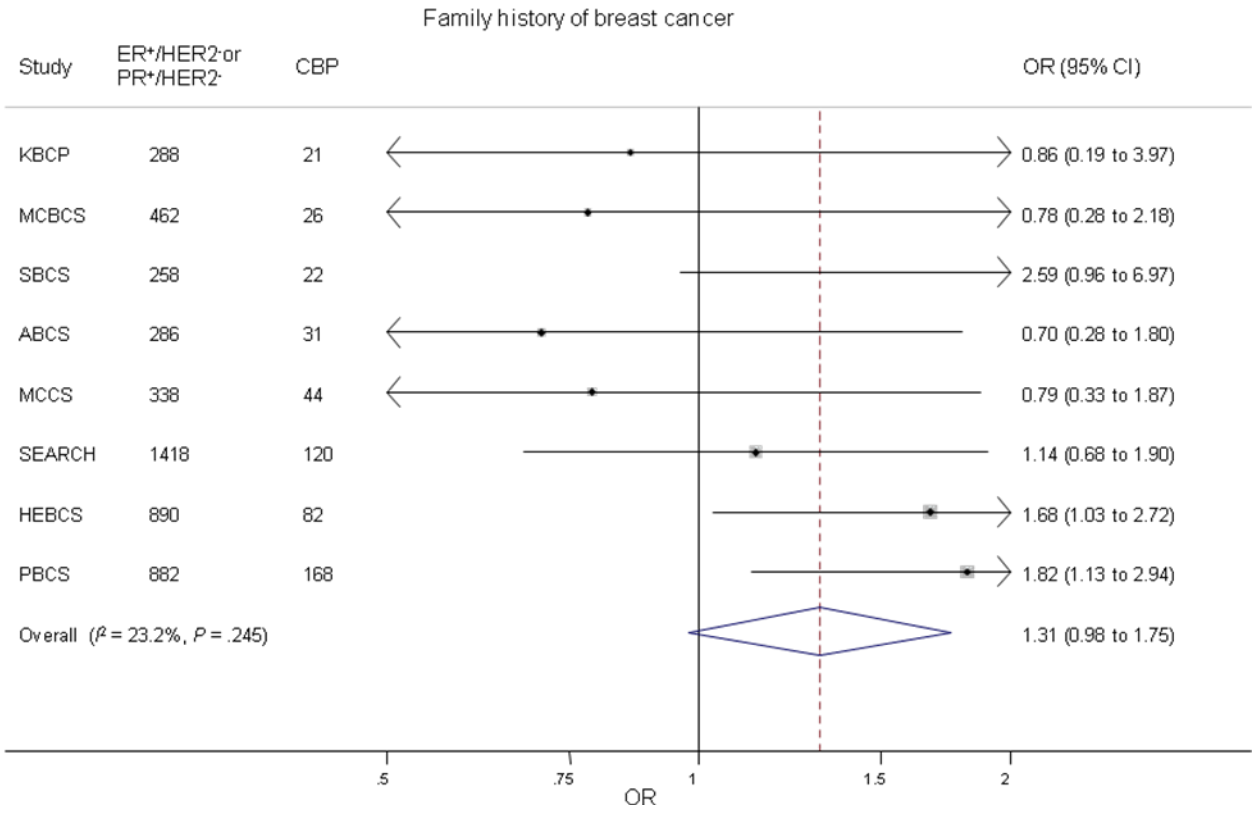
B.


Nulliparity
Study
ER+
ER-
OR (95% CI)
kConFab/AOCS 127 63 0.40 (0.11 to 1.50)
ORIGO 547 198 0.68 (0.22 to 2.10)
CNIO-BCS 102 32 0.65 (0.26 to 1.63)
OFBCR 534 212 0.94 (0.39 to 2.26)
OBCS 364 91 1.07 (0.55 to 2.08)
KARBAC 373 78 1.46 (0.77 to 2.78)
HABCS 565 73 0.81 (0.43 to 1.52)
NHS 707 186 1.21 (0.67 to 2.19)
KBCP 336 101 1.03 (0.58 to 1.84)
TBCS 117 126 0.85 (0.48 to 1.49)
LMBC 554 126 0.92 (0.53 to 1.60)
MEC 658 123 0.86 (0.50 to 1.46)
RBCS 409 137 1.03 (0.62 to 1.71)
MCBCS 867 175 1.08 (0.66 to 1.77)
ABCFS 724 407 0.89 (0.55 to 1.46)
TWBCS 460 271 0.92 (0.57 to 1.49)
SBCS 524 175 0.96 (0.59 to 1.55)
SASBAC 831 225 0.63 (0.39 to 1.01)
UCIBCS 594 158 1.14 (0.71 to 1.85)
CGPS 1145 220 1.03 (0.65 to 1.62)
MCCS 640 238 0.45 (0.29 to 0.70)
BBCC 745 277 0.69 (0.45 to 1.07)
GESBC 314 207 1.07 (0.70 to 1.64)
GENICA 755 216 0.95 (0.64 to 1.42)
NC-BCFR 917 309 0.65 (0.46 to 0.91)
HEBCS 1341 306 0.80 (0.59 to 1.09)
PBCS 1189 615 0.78 (0.58 to 1.03)
MARIE 2010 537 0.74 (0.57 to 0.97)
SEARCH 3540 856 0.79 (0.63 to 1.00)
Overall (I^2 = 0.0%, P = .538) 0.83 (0.77 to 0.90)
.1
.5
1
1.5
2
OR

C.


Age at first full-term birth
Study
ER+
ER-
OR (95% CI)
LMBC 152 36 1.42 (0.92 to 2.17)
kConFab/AOCS 113 59 0.79 (0.56 to 1.11)
TBCS 80 90 0.87 (0.64 to 1.19)
KBCP 268 81 0.96 (0.73 to 1.27)
KARBAC 281 58 1.09 (0.83 to 1.44)
RBCS 313 97 0.84 (0.64 to 1.10)
GESBC 245 159 0.74 (0.60 to 0.93)
SBCS 433 145 1.00 (0.81 to 1.23)
TWBCS 404 241 0.95 (0.78 to 1.16)
UCIBCS 499 130 1.04 (0.86 to 1.26)
OFBCR 512 201 0.87 (0.71 to 1.05)
ORIGO 532 194 0.91 (0.75 to 1.10)
GENICA 622 175 0.95 (0.79 to 1.15)
MCBCS 750 145 0.92 (0.77 to 1.11)
MCCS 505 209 0.93 (0.78 to 1.12)
SASBAC 704 201 0.85 (0.72 to 1.01)
BBCC 585 232 1.04 (0.89 to 1.22)
HEBCS 957 225 0.95 (0.82 to 1.10)
ABCFS 677 378 0.88 (0.77 to 1.02)
NC-BCFR 705 255 0.86 (0.76 to 0.98)
PBCS 1001 537 0.82 (0.72 to 0.92)
MARIE 1657 462 0.92 (0.82 to 1.03)
SEARCH 2955 720 0.87 (0.79 to 0.95)
Overall (I^2 = 1.8%, P = .436) 0.90 (0.87 to 0.94)
.5
.75
1
1.25
1.5
OR

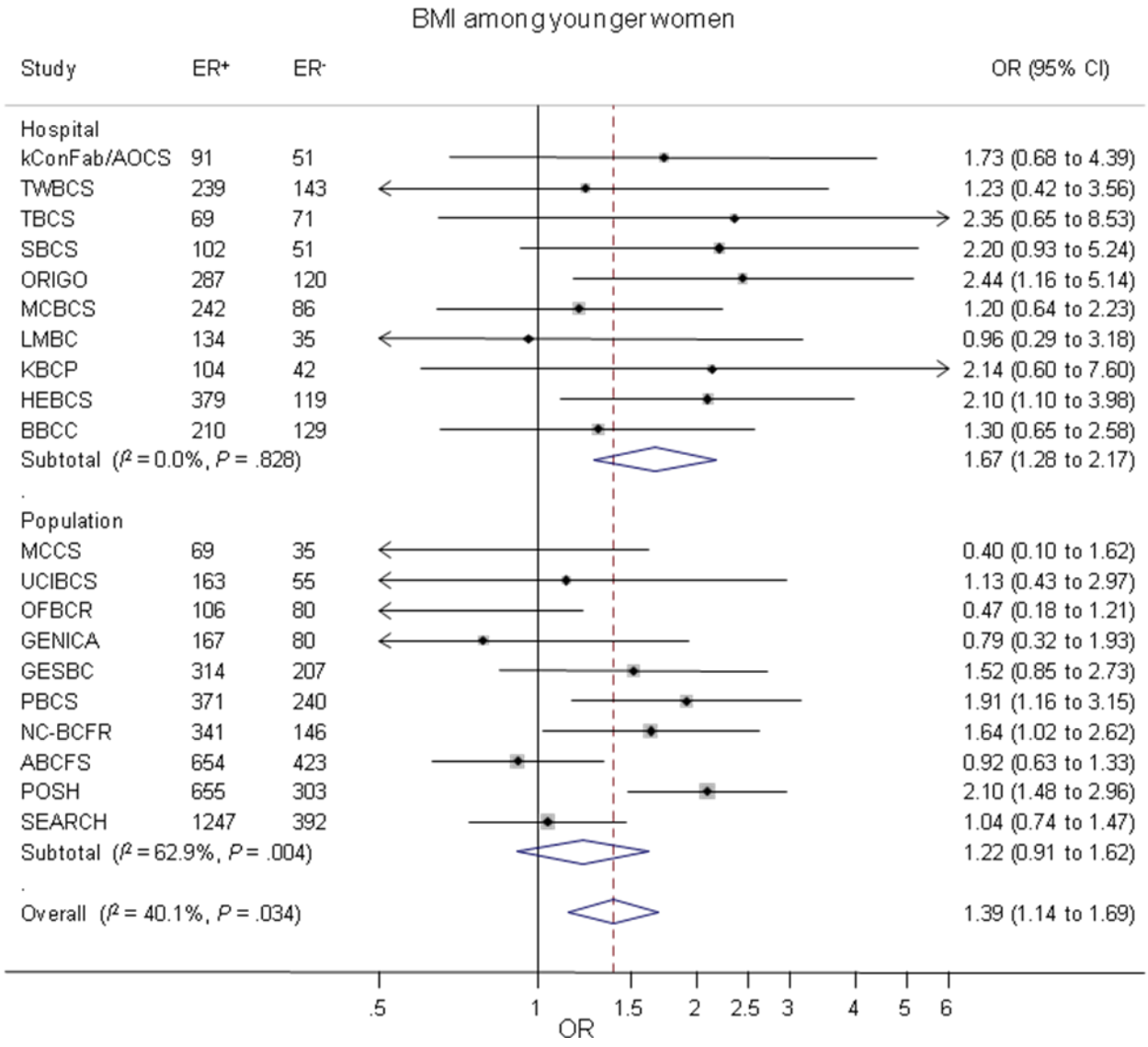
D.


Family history of breast cancer
Study
ER+/HER2-or PR+/HER2-
CBP
OR (95% CI)
KBCP 288 21 0.86 (0.19 to 3.97)
MCBCS 462 26 0.78 (0.28 to 2.18)
SBCS 258 22 2.59 (0.96 to 6.97)
ABCS 286 31 0.70 (0.28 to 1.80)
MCCS 338 44 0.79 (0.33 to 1.87)
SEARCH 1418 120 1.14 (0.68 to 1.90)
HEBCS 890 82 1.68 (1.03 to 2.72)
PBCS 882 168 1.82 (1.13 to 2.94)
Overall (I^2 = 23.2%, P = .245) 1.31 (0.98 to 1.75)
.5
.75
1
1.5
2
OR

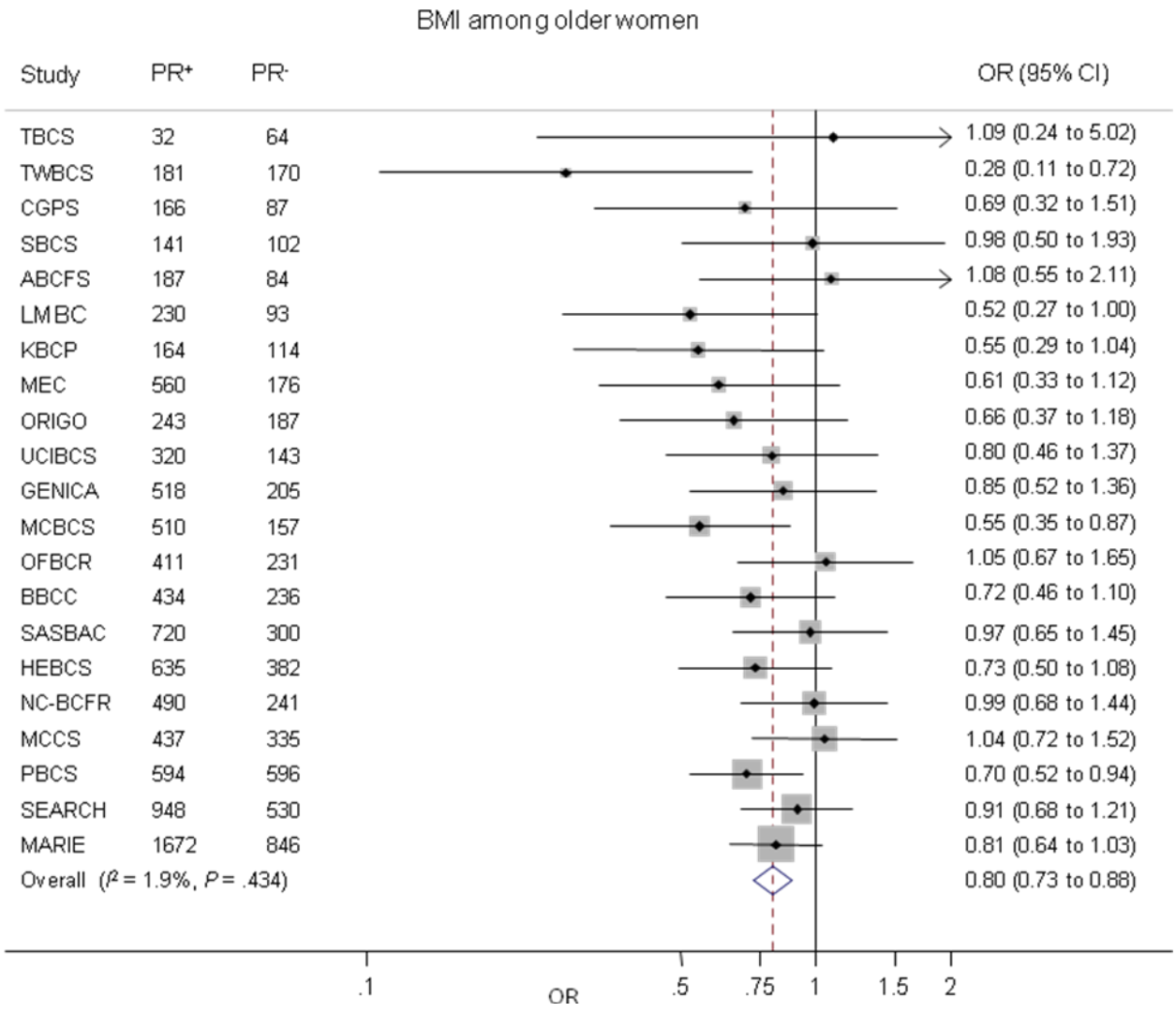
Supplementary Figure 1. Study–specific case–case odds ratios (ORs) and 95% confidence intervals (CIs) for associations between reproductive risk factors and tumor subtypes defined by marker expression. Pooled effect of each risk factor was derived from the point estimate for each study weighted by the inverse of the variance. Summary ORs and 95% CIs were estimated using a random–effects model of DerSimonian and Laird. $P<0.05$ was used to define statistically significant heterogeneity across studies. All statistical tests were two–sided. Studies were sorted by percent weight of each study contributing to the summary OR. The dot on each square represents the study–specific OR, and the size of the square represents the weight of each study. The horizontal lines represent the CIs; if ending in an arrow, this indicates that the interval transcends the region plotted. The diamond represents the summary OR. Solid vertical lines represent an OR of 1; dashed vertical lines represent the overall ORs. **A)** Case–case ORs for the risk of having PR–negative tumors (comparing to PR–positive tumors) for women with younger age at menarche (≤12 years) compared to women with age at menarche ≥15 years by study. ORs and 95% CIs were obtained from unconditional logistic regression models with PR status as the outcome variable and age at menarche (comparing age ≤12 to ≥15 years) and age at diagnosis as independent variables. **B)** Case–case ORs for the risk of having ER–negative tumors (comparing to ER–positive tumors) for nulliparous women compared to parous women by study. ORs and 95% CIs were obtained from unconditional logistic regression models with ER status as the outcome variable and parity and age at diagnosis as independent variables. The reference group is parous women. **C)** Case–case ORs for the risk of having ER–negative tumors (comparing to ER–positive tumors) for parous women with a 5-year increasing age at first full term birth by study. ORs and 95% CIs were obtained from unconditional logistic regression models with ER status as the outcome variable and age at first full term birth (continuous, per 5-year increase) and age at diagnosis as independent variables. **D)** Case–case ORs for the risk of core basal phenotype (CBP) tumors comparing to $ER^+/HER2^-$ or $PR^+/HER2^-$tumors associated with positive family history of breast cancer by study. ORs and 95% CIs were obtained from unconditional logistic regression models with tumor subtype as the outcome variable (comparing CBP to $ER^+/HER2^-$ or $PR^+/HER2^-$) and family history and age at diagnosis as independent variables.

Supplementary Figure 2

A.



B.


BMI among older women
Study
PR+
PR-
OR (95% CI)
TBCS 32 64 1.09 (0.24 to 5.02)
TWBCS 181 170 0.28 (0.11 to 0.72)
CGPS 166 87 0.69 (0.32 to 1.51)
SBCS 141 102 0.98 (0.50 to 1.93)
ABCFS 187 84 1.08 (0.55 to 2.11)
LMBC 230 93 0.52 (0.27 to 1.00)
KBCP 164 114 0.55 (0.29 to 1.04)
MEC 560 176 0.61 (0.33 to 1.12)
ORIGO 243 187 0.66 (0.37 to 1.18)
UCIBCS 320 143 0.80 (0.46 to 1.37)
GENICA 518 205 0.85 (0.52 to 1.36)
MCBCS 510 157 0.55 (0.35 to 0.87)
OFBCR 411 231 1.05 (0.67 to 1.65)
BBCC 434 236 0.72 (0.46 to 1.10)
SASBAC 720 300 0.97 (0.65 to 1.45)
HEBCS 635 382 0.73 (0.50 to 1.08)
NC-BCFR 490 241 0.99 (0.68 to 1.44)
MCCS 437 335 1.04 (0.72 to 1.52)
PBCS 594 596 0.70 (0.52 to 0.94)
SEARCH 948 530 0.91 (0.68 to 1.21)
MARIE 1672 846 0.81 (0.64 to 1.03)
Overall (I^2 = 1.9%, P = .434) 0.80 (0.73 to 0.88)
.1
.5
.75
1
1.5
2
OR

Supplementary Figure 2. Study–specific case–case odds ratios (ORs) and 95% confidence intervals (CIs) for associations between reproductive risk factors and tumor subtypes defined by marker expression. Pooled effect of each risk factor was derived from the point estimate for each study weighted by the inverse of the variance. Summary ORs and 95% CIs were estimated using a random–effects model of DerSimonian and Laird. $P<0.05$ was used to define significant heterogeneity across studies. All statistical tests were two–sided. Studies were sorted by % weight of each study contributing to the summary OR. The dot on each square represents the study–specific OR, and the size of the square represents the weight of each study. The horizontal lines represent the CIs; if ending in an arrow, this indicates that the interval transcends the region plotted. The diamond represents the summary OR. Solid vertical lines represent an OR of 1; dashed vertical lines represent the overall ORs. NOTE: Definition of study design in this figure is based only on the source of case patients (not the source of control subjects), which differs in some studies from the study design definition for case–control analyses which is based on the source of case patients and control subjects. **A)** Case–case ORs for the risk of having ER–negative tumors (comparing to ER–positive tumors) associated with BMI among younger case patients (age≤50) by study (grouped by study design). ORs and 95% CIs were obtained from unconditional logistic regression models with ER status as the outcome variable and BMI among younger case patients (comparing BMI≥30 to BMI<25) and age at diagnosis as independent variables. **B)** Case–case ORs for the risk of having PR–negative tumors (comparing to PR–positive tumors) associated with BMI among older case patients (age>50) by study. Ors and 95% CIs were obtained from unconditional logistic regression models with PR status as the outcome variable and BMI among older case patients (comparing BMI≥30 to BMI<25) and age at diagnosis as independent variables.

References

1. Dite GS, Jenkins MA, Southey MC, *et al.* Familial risks, early-onset breast cancer, and BRCA1 and BRCA2 germline mutations. J Natl Cancer Inst 2003;95(6):448-57.
2. Schmidt MK, Tollenaar RA, de Kemp SR, *et al.* Breast cancer survival and tumor characteristics in premenopausal women carrying the CHEK2*1100delC germline mutation. J Clin Oncol 2007;25(1):64-9.
3. Fasching PA, Loehberg CR, Strissel PL, *et al.* Single nucleotide polymorphisms of the aromatase gene (CYP19A1), HER2/neu status, and prognosis in breast cancer patients. Breast Cancer Res Treat 2008;112(1):89-98.
4. Schrauder M, Frank S, Strissel PL, *et al.* Single nucleotide polymorphism D1853N of the ATM gene may alter the risk for breast cancer. J Cancer Res Clin Oncol 2008;134(8):873-82.
5. Colleran G, McInerney N, Rowan A, *et al.* The TGFBR1*6A/9A polymorphism is not associated with differential risk of breast cancer. Breast Cancer Res Treat 2009.
6. McInerney N, Colleran G, Rowan A, *et al.* Low penetrance breast cancer predisposition SNPs are site specific. Breast Cancer Res Treat 2009;117(1):151-9.
7. Bojesen SE, Tybjaerg-Hansen A, Axelsson CK, *et al.* No association of breast cancer risk with integrin beta3 (ITGB3) Leu33Pro genotype. Br J Cancer 2005;93(1):167-71.
8. Weischer M, Bojesen SE, Tybjaerg-Hansen A, *et al.* Increased risk of breast cancer associated with CHEK2*1100delC. J Clin Oncol 2007;25(1):57-63.
9. Milne RL, Ribas G, Gonzalez-Neira A, *et al.* ERCC4 associated with breast cancer risk: a two-stage case-control study using high-throughput genotyping. Cancer Res 2006;66(19):9420-7.
10. Pesch B, Ko Y, Brauch H, *et al.* Factors modifying the association between hormone-replacement therapy and breast cancer risk. Eur J Epidemiol 2005;20(8):699-711.
11. Justenhoven C, Pierl CB, Haas S, *et al.* The CYP1B1_1358_GG genotype is associated with estrogen receptor-negative breast cancer. Breast Cancer Res Treat 2008;111(1):171-7.
12. Chang-Claude J, Eby N, Kiechle M, *et al.* Breastfeeding and breast cancer risk by age 50 among women in Germany. Cancer Causes Control 2000;11(8):687-95.
13. Dork T, Bendix R, Bremer M, *et al.* Spectrum of ATM gene mutations in a hospital-based series of unselected breast cancer patients. Cancer Res 2001;61(20):7608-15.
14. Syrjakoski K, Vahteristo P, Eerola H, *et al.* Population-based study of BRCA1 and BRCA2 mutations in 1035 unselected Finnish breast cancer patients. J Natl Cancer Inst 2000;92(18):1529-31.
15. Kilpivaara O, Bartkova J, Eerola H, *et al.* Correlation of CHEK2 protein expression and c.1100delC mutation status with tumor characteristics among unselected breast cancer patients. Int J Cancer 2005;113(4):575-80.
16. Lindblom A, Rotstein S, Larsson C, *et al.* Hereditary breast cancer in Sweden: a predominance of maternally inherited case patients. Breast Cancer Res Treat 1992;24(2):159-65.
17. Margolin S, Werelius B, Fornander T, *et al.* BRCA1 mutations in a population-based study of breast cancer in Stockholm County. Genet Test 2004;8(2):127-32.
18. Hartikainen JM, Tuhkanen H, Kataja V, *et al.* An autosome-wide scan for linkage disequilibrium-based association in sporadic breast cancer case patients in eastern Finland: three candidate regions found. Cancer Epidemiol Biomarkers Prev 2005;14(1):75-80.
19. Hartikainen JM, Tuhkanen H, Kataja V, *et al.* Refinement of the 22q12-q13 breast cancer--associated region: evidence of TMPRSS6 as a candidate gene in an eastern Finnish population. Clin Cancer Res 2006;12(5):1454-62.

20. Beesley J, Jordan SJ, Spurdle AB, *et al.* Association between single-nucleotide polymorphisms in hormone metabolism and DNA repair genes and epithelial ovarian cancer: results from two Australian studies and an additional validation set. Cancer Epidemiol Biomarkers Prev 2007;16(12):2557-65.
21. Mann GJ, Thorne H, Balleine RL, *et al.* Analysis of cancer risk and BRCA1 and BRCA2 mutation prevalence in the kConFab familial breast cancer resource. Breast Cancer Res 2006;8(1):R12.
22. De Maeyer L, Van Limbergen E, De Nys K, *et al.* Does estrogen receptor negative/progesterone receptor positive breast carcinoma exist? J Clin Oncol 2008;26(2):335-6; author reply 336-8.
23. Neven P, Brouckaert O, Van Belle V, *et al.* In early-stage breast cancer, the estrogen receptor interacts with correlation between human epidermal growth factor receptor 2 status and age at diagnosis, tumor grade, and lymph node involvement. J Clin Oncol 2008;26(10):1768-9; author reply 1769-71.
24. Flesch-Janys D, Slanger T, Mutschelknauss E, *et al.* Risk of different histological types of postmenopausal breast cancer by type and regimen of menopausal hormone therapy. Int J Cancer 2008;123(4):933-41.
25. Catucci I, Verderio P, Pizzamiglio S, *et al.* SNPs in ultraconserved elements and familial breast cancer risk. Carcinogenesis 2009;30(3):544-5; author reply 546.
26. De Vecchi G, Verderio P, Pizzamiglio S, *et al.* Evidences for association of the CASP8 -652 6N del promoter polymorphism with age at diagnosis in familial breast cancer case patients. Breast Cancer Res Treat 2009;113(3):607-8.
27. Olson JE, Ma CX, Pelleymounter LL, *et al.* A comprehensive examination of CYP19 variation and breast density. Cancer Epidemiol Biomarkers Prev 2007;16(3):623-5.
28. Giles GG, English DR. The Melbourne Collaborative Cohort Study. IARC Sci Publ 2002;156:69-70.
29. Kolonel LN, Henderson BE, Hankin JH, *et al.* A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol 2000;151(4):346-57.
30. John EM, Hopper JL, Beck JC, *et al.* The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. Breast Cancer Res 2004;6(4):R375-89.
31. Hankinson SE, Willett WC, Manson JE, *et al.* Plasma sex steroid hormone levels and risk of breast cancer in postmenopausal women. J Natl Cancer Inst 1998;90(17):1292-9.
32. Thomas G, Jacobs KB, Kraft P, *et al.* A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat Genet 2009;41(5):579-84.
33. Erkko H, Xia B, Nikkila J, *et al.* A recurrent mutation in PALB2 in Finnish cancer families. Nature 2007;446(7133):316-9.
34. de Bock GH, Schutte M, Krol-Warmerdam EM, *et al.* Tumour characteristics and prognosis of breast cancer patients carrying the germline CHEK2*1100delC variant. J Med Genet 2004;41(10):731-5.
35. Huijts PE, Vreeswijk MP, Kroeze-Jansema KH, *et al.* Clinical correlates of low-risk variants in FGFR2, TNRC9, MAP3K1, LSP1 and 8q24 in a Dutch cohort of incident breast cancer case patients. Breast Cancer Res 2007;9(6):R78.
36. Garcia-Closas M, Egan KM, Newcomb PA, *et al.* Polymorphisms in DNA double-strand break repair genes and risk of breast cancer: two population-based studies in USA and Poland, and meta-analyses. Hum Genet 2006;119(4):376-88.

37. Eccles D, Gerty S, Simmonds P, *et al.* Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH): study protocol. BMC Cancer 2007;7:160.
38. Tapper W, Hammond V, Gerty S, *et al.* The influence of genetic variation in 30 selected genes on the clinical characteristics of early onset breast cancer. Breast Cancer Res 2008;10(6):R108.
39. Easton DF, Pooley KA, Dunning AM, *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447(7148):1087-93.
40. Wedren S, Lovmar L, Humphreys K, *et al.* Oestrogen receptor alpha gene haplotype and postmenopausal breast cancer risk: a case control study. Breast Cancer Res 2004;6(4):R437-49.
41. MacPherson G, Healey CS, Teare MD, *et al.* Association of a common variant of the CASP8 gene with reduced risk of breast cancer. J Natl Cancer Inst 2004;96(24):1866-9.
42. Rafii S, O'Regan P, Xinarianos G, *et al.* A potential role for the XRCC2 R188H polymorphic site in DNA-damage repair and breast cancer. Hum Mol Genet 2002;11(12):1433-8.
43. Lesueur F, Pharoah PD, Laing S, *et al.* Allelic association of the human homologue of the mouse modifier Ptprj with breast cancer. Hum Mol Genet 2005;14(16):2349-56.
44. Jakubowska A, Jaworska K, Cybulski C, *et al.* Do BRCA1 modifiers also affect the risk of breast cancer in non-carriers? Eur J Cancer 2009;45(5):837-42.
45. Lubinski J, Korzen M, Gorski B, *et al.* Genetic contribution to all cancers: the first demonstration using the model of breast cancers from Poland stratified by age at diagnosis and tumour pathology. Breast Cancer Res Treat 2009;114(1):121-6.
46. Sangrajrang S, Schmezer P, Burkholder I, *et al.* Polymorphisms in three base excision repair genes and breast cancer risk in Thai women. Breast Cancer Res Treat 2008;111(2):279-88.
47. Ding SL, Yu JC, Chen ST, *et al.* Genetic variants of BLM interact with RAD51 to increase breast cancer susceptibility. Carcinogenesis 2009;30(1):43-9.
48. Hsu HM, Wang HC, Chen ST, *et al.* Breast cancer risk is associated with the genes encoding the DNA double-strand break repair Mre11/Rad50/Nbs1 complex. Cancer Epidemiol Biomarkers Prev 2007;16(10):2024-32.
49. Anton-Culver H, Cohen PF, Gildea ME, *et al.* Characteristics of BRCA1 mutations in a population-based case series of breast and ovarian cancer. Eur J Cancer 2000;36(10):1200-8.
50. Ziogas A, Gildea M, Cohen P, *et al.* Cancer risk estimates for family members of a population-based family registry for breast and ovarian cancer. Cancer Epidemiol Biomarkers Prev 2000;9(1):103-11.